(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 630 972 A3**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **94109745.3**

(51) Int. Cl.[6] **C12Q 1/68**

(22) Date of filing: **23.06.94**

(30) Priority: **25.06.93 JP 155534/93**
30.07.93 JP 189624/93

(43) Date of publication of application:
28.12.94 Bulletin 94/52

(84) Designated Contracting States:
DE GB

(88) Date of deferred publication of the search report:
29.11.95 Bulletin 95/48

(71) Applicant: **HITACHI, LTD.**
6, Kanda Surugadai 4-chome
Chiyoda-ku,
Tokyo 100 (JP)

(72) Inventor: **Kambara, Hideki**
1-4-3, Kitanodai
Hachiouji-shi,
Tokyo 192 (JP)
Inventor: **Okano, Kazunori**
5-17-2-402, Hon-cho
Shiki-shi,
Saitama-ken 353 (JP)
Inventor: **Takahashi, Satoshi**
Nr. 202, Dai-12 Shin'ei-manshon,
1-5-17, Nishi
Kunitachi-shi,
Tokyo 186 (JP)
Inventor: **Nagai, Kelichi**
3-44-14, Sakuragaoka
Higashiyamato-shi,
Tokyo 207 (JP)
Inventor: **Kawamoto, Kazuko, Nr. 309, Hitachi**
Ozaki-haitsu
3-8-9, Higashi-koigakubo
Kokubunji-shi,
Tokyo 185 (JP)
Inventor: **Furuyama, Hiroko, Nr. 110, Hitachi**
Ozaki-haitsu
3-8-9, Higashi-koigakubo
Kokubunji-shi,
Tokyo 185 (JP)

(74) Representative: **Strehl Schübel-Hopf Groening & Partner**
Maximilianstrasse 54
D-80538 München (DE)

(54) DNA analyzing method.

(57) A DNA analyzing method which bonds the oligomer of the known base sequence to the DNA fragment obtained by digesting sample 1 with restrictive enzyme, causes hybridization between oligomer 3 and DNA fragment using the oligomers 3 which have the sequences of all combinations of the types of the bases within the length of several bases following the known base sequence, checks presence or absence of the hybridization or complementary DNA strand extension, identifies the DNA fragment terminal sequence from this result, and fractionates the DNA fragments and analyzes them or analyzes them as they are. This DNA analyzing method provides an effective analysis of mixtures of long DNAs or DNA fragments.

EP 0 630 972 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 10 9745

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 86, 1989 WASHINGTON US, pages 8902-8906, S.BRENNER ET AL. 'DNA fingerprinting by sampled sequencing'	1,13	C12Q1/68
Y	* the whole document *	2,12	
X	EP-A-0 309 969 (E.I. DUPONT DE NEMOURS AND COMPANY) 5 April 1989 * the whole document *	1	
Y	EP-A-0 303 459 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 15 February 1989	2,12	
A	* the whole document *	5	
A	WO-A-89 10977 (ISIS INNOVATION LIMITED) 16 November 1989 * the whole document, esp. page 24 *	1,3,8,9, 11,12	
A	EP-A-0 224 126 (THE UNIVERSITY OF CALGARY) 3 June 1987 * the whole document *	2,5,12	TECHNICAL FIELDS SEARCHED (Int.Cl.5) C12Q
A	WO-A-92 10588 (AFFYMAX TECHNOLOGIES N.V.) 25 June 1992 * the whole document *	1-13	
D,A	SCIENCE, vol. 251, no. 4995, 15 February 1991 pages 767-773, XP 000486899 S.P.A. FODOR ET AL. 'Light-directed, spatially addressable parallel chemical synthesis' * the whole document *	1-13	

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	20 September 1995	De Kok, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C09)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 10 9745

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
P,X	WO-A-93 17126 (THE PUBLIC HEALTH RESEARCH INSTITUTE OF THE CITY OF NEW YORK, INC.) 2 September 1993 * page 8, line 21 - page 13, line 17 * * page 17, last paragraph - page 30, line 16 * ---	1-13	
T	BIOTECHNOLOGY, vol. 12, November 1994 NEW YORK US, pages 1093-1099, A.B. CHETVERIN ET AL. 'Oligonucleotide arrays: new concepts and possibilities' * the whole document * -----	1-13	

TECHNICAL FIELDS SEARCHED (Int.Cl.5)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	20 September 1995	De Kok, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 630 972 A2

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 94109745.3

(51) Int. Cl.5 C12Q 1/68

(22) Date of filing: 23.06.94

(30) Priority: 25.06.93 JP 155534/93
30.07.93 JP 189624/93

(43) Date of publication of application:
28.12.94 Bulletin 94/52

(84) Designated Contracting States:
DE GB

(71) Applicant: HITACHI, LTD.
6, Kanda Surugadai 4-chome
Chiyoda-ku,
Tokyo 100 (JP)

(72) Inventor: Kambara, Hideki
1-4-3, Kitanodai
Hachiouji-shi,
Tokyo 192 (JP)
Inventor: Okano, Kazunori
5-17-2-402, Hon-cho
Shiki-shi,
Saitama-ken 353 (JP)
Inventor: Takahashi, Satoshi
Nr. 202, Dai-12 Shin'ei-manshon,
1-5-17, Nishi
Kunitachi-shi,
Tokyo 186 (JP)
Inventor: Nagai, Keiichi
3-44-14, Sakuragaoka
Higashiyamato-shi,
Tokyo 207 (JP)
Inventor: Kawamoto, Kazuko, Nr. 309, Hitachi
Ozaki-haitsu
3-8-9, Higashi-koigakubo
Kokubunji-shi,
Tokyo 185 (JP)
Inventor: Furuyama, Hiroko, Nr. 110, Hitachi
Ozaki-haitsu
3-8-9, Higashi-koigakubo
Kokubunji-shi,
Tokyo 185 (JP)

(74) Representative: Strehl Schübel-Hopf Groening
& Partner
Maximilianstrasse 54
D-80538 München (DE)

(54) DNA analyzing method.

(57) A DNA analyzing method which bonds the oligomer of the known base sequence to the DNA fragment obtained by digesting sample 1 with restrictive enzyme, causes hybridization between oligomer 3 and DNA fragment using the oligomers 3 which have the sequences of all combinations of the types of the bases within the length of several bases following the known base sequence, checks presence or absence of the hybridization or complementary DNA strand extension, identifies the DNA fragment terminal sequence from this result, and fractionates the DNA fragments and analyzes them or analyzes them as they are. This DNA analyzing method provides an effective analysis of mixtures of long DNAs or DNA fragments.

EP 0 630 972 A2

BACKGROUND OF THE INVENTION

The present invention relates to the DNA analyzing method including DNA sequencing.

Conventional DNA sequencing of large DNA (e.g. 10^5 to 10^6 (100 K to 1 M) base length) requires subcloning processes. In the subcloning processes, the target DNA is digested with restriction enzymes to produce fragments of the target DNA. The fragments are cloned in a plasmid, which is used as a vector in the cloning process. E. Coli infected with the plasmid is cultured in the agar thereafter. As only infected E. Coli can survive and make colony in the agar, we can select E. Coli infected with the plasmid by picking up the colonies. Then the target DNA fragments subcloned in the plasmid are extracted from E. Coli. The extracted target DNA fragments are sequenced. Then the target DNA fragments are taken separately by picking up obtained colonies, respectively. DNA sequencing has been carried out for each colony using the plasmid containing the DNA fragment taken separately. Normally, 300 to 500 DNA bases can be determined by one sequencing operation, and sequencing of 1 mega (10^6) bases requires analysis of 2000 or more colonies. Furthermore, the colonies containing same DNA fragments may be picked up for the analysis; this requires the colonies in the number of several times 2,000, namely, close to 10,000 colonies to be picked up for analysis.

According to the conventional method, base sequencing for human genome and other creatures, genome requires digesting of a great length of DNA by restriction enzymes and subcloning of the digested DNA into vectors such as plasmid and yeast clomosomes, thereby forming a colony derived from one containing one DNA digested product and selecting each DNA fragment by picking up the colony. Then DNA fragments have been amplified by culturing each of such microbe obtained from one colony to obtain a great amount of the DNA fragment required for base sequencing operation. Called "cloning or subcloning", this biological technique has disadvantages of taking a lot of time for culturing and is unfitted for automation. Such conventional techniques are discussed in Molecular Cloning, A Laboratory Manual (2nd Version) (Chapters 1 to 4 and Chapter 9, Cold spring Harbor Laboratory Press, 1989).

According to the above technique, single DNA fragment species required for sequencing operation has been prepared biologically using vectors and infected colibacillus and yeast. This method requires the use of a special facility called P2. Another disadvantage of this method is to take a long time to culture the microbes and picking up a large number of colonies for analysis, which are unfit for automation.

Primer Walking (Science, Vol. 258, pages 1787 to 1791, 1992) is known as a method for long DNA analysis; it provides sequencing of 40 DNA base from one end using the first primer. The second sequencing step is carried out with the second primer which is the oligomer having the sequence in the vicinity of the first Sequenced DNA terminal. The sequencing is carried out step by step using new promers which has the sequences complementary to the sequenced DNA. Analysis of the 1 mega base DNA requires 2500 sequencing steps with this method and each analysis requires one or more days; thus several years have been required for all analyses.

Thus, such a large scale DNA sequencing method comprises processes requiring much cost and time. Depending on the sizes and types of the DNA fragments digested by restrictive enzyme, some of the DNA fragments have been difficult to be introduced into the vector by cloning. Consequently, a large scale DNA sequencing has been a problem requiring solutions by developing of a new technique.

SUMMARY OF THE INVENTION

The object of the present invention is to provide solutions to the problems of said conventional techniques and a DNA analysis method ensuring effective analysis of long DNAs or mixtures of DNA fragments. Another object of the present invention is to provide a method which allows DNA sequencing without cloning and permits automatic amplification and selection of DNAs.

To achieve said objectives, the DNA analysis according to the present invention includes:

(1) a process of digesting the long DNA at a certain sequence portion using an enzyme and the like, and ligating a DNA, which is labeled with a fluorophore or the like, and has a known sequence, to the digested portion,

(2) a process of forming a single stranded DNA at least in the vicinity of the terminal of the double strand,

(3) a process of separating DNA according to the difference of the terminal sequences following to the ligated known oligomer sequence and,

(4) a process of fractionating and sequencing the separated DNA.

The present invention provides a DNA analyzing method characterized by including a process of introducing a known sequence into the terminal of the DNA fragment by ligating oligomer, and a process of sequencing the specific DNA fragment alone which has a certain base sequences (2 mer to 6 mer) following to the ligated oligomer DNA sequence at the terminal of this DNA fragment.

The DNA is digested by a restriction enzyme having the property of recognizing and digesting a certain sequence. The restriction enzyme used in the present case can be any enzyme; Hind III and Not I are used for example. The terminal base of the said DNA fragment can be labeled with fluorophore or the like to ensure that many types of the generated DNA fragments can be easily separated and taken out (Fig. 2). Fluorophores such as FITC (fluoresceine isothiocyanate having an emission wavelength of 525 nm) and Texas Red (sulforhodamine 101 having an emission wavelength of 613 nm), chemical luminescence or RI are used as label. Many fragment species are produced through the digestion. As too many DNA species is not good for amplification of DNA fragments by PCR, the DNA fragments are separated by gel electrophoresis to be fractionated before the PCR amplification. This separation may not be very strict and a fraction can contain several DNA fragment species. In addition to gel electrophoresis, DNA probe array, liquid chromatograph or affinity chromatograph can be used as the separation means. When the amount of DNA fragments is not sufficient for DNA sequencing, the number of DNA fragment copies can be amplified by repeated DNA polymerase reactions (PCR amplification) using the ligated known sequences at the both ends of DNA fragments.

The DNA fragments obtained in this stage are composed of several species of DNA fragments; one specific DNA fragment species is in a more strict manner to determine the base sequence. Said selective base sequencing will be done in the following procedure:

The oligopolynucleotide having a base sequence comprising the known DNA sequence complementary to the oligopolynucleotide ligated to the portion digested by the restriction enzyme and additional 2- to 6-base sequence at its 3' terminal, which hybridizes on the ligated oligomer and 2- to 6- base sequence following the digested portion of DNA fragments is added to the mixture of several species of DNA fragments as primer, so that the primer will hybridize on the 3' terminal of the specific DNA fragments. Only the specific template DNA is selected to undergo complementary DNA strand extension in the presence of four bases of adenine (A), thymine (T), guanine (G), cytosine (C), ddATP, ddTTP, ddGTP, and ddCTP, and DNA polymerase; then the sequencing reaction products are used for selective quenching. Each DNA fragment group where 3' terminals of the hybridized strand obtained comprise adenine, thymine, guanine, cytosine is analyzed by electrophoresis to determine DNA fragment length and base sequence. Furthermore, to enhance selectivity of the said primer, complementary DNA strand extension can be carried out at the temperature of 60 degrees Celsius or more where primer bonding performance is reduced and only perfectly hybridized primers can work. This is because hybridization between said primer and DNA fragments is provided by hydrogen bonding, so when reaction temperature is increased, stability is maintained at the 3' terminal sequence portion of the primer only for the fragments having the sequence where hybridization exhibits a complete agreement, and, in other fragments, complementary DNA strand extension does not take place over that temperature. Therefore, when hybridization reaction temperature (annealing temperature) is raised to 60 degrees Celsius or more, more effective selection will be ensured, thereby providing effective selectivity.

The following discusses the principle of said selective DNA sequencing according to Figs. 6, 10A and 10B. In the DNA sequencing operation in the final process of Fig. 6, many kinds of template DNAs 109 or 111 and one type of primer 110 are mixed to cause hybridization, and only the template DNAs 109 capable of complete hybridization with primer 110 are sequenced. In greater details, DNA fragment group 107 comprising of five to six kinds in Fig. 6 is the double stranded DNA comprising the known DNA sequence portion connected to the restrictive enzyme digested portion at both terminals and the unknown DNA sequence portion at the center. These DNA fragments are heated and changed into single stranded DNA 109 and 111 in Figs. 10A and 10B, and are hybridized with the primer 110 comprising the sequence complementary to the known sequence portion at the terminus and the base sequence complementary to two bases at the 5' terminal of the unknown sequence portion (portion where two bases at the 3' terminal of the primer will be bonded) following it. Two bases on the terminal side of DNA fragment followed by the known sequence portion of ligated oligomers 109 and 111 are unknown bases, and the sequence differs according to the species of the DNA fragments. Therefore, the type of the primer capable of complete hybridization differs depending on the species of the DNA. This example is illustrated by a case of primer having the known sequence and additional two bases, however the additional base part can be increased up to 6 bases. The number of primer types depends on the additional base number. It is 16 for two bases and 4^6 for six bases. When the number of template DNA (or DNA fragment to be analyzed) species is much less than the number of primer types, the possibility of one primer hybridization on more than two template DNA species is very low. This means that only one type of DNA fragments (Fig. 10A) is subjected to hybridization for one type of primers, so only one type of DNA sequence can

be read out, analyzed and determined on a selective basis. It should be noted, however, that, the length of the primer, which hybridizes to the known DNA sequence portion at the 5' terminal and the additional several bases following it, is more than six bases for the primer to function. The primer longer than 16 mer is preferred from the viewpoint of stability. The length of primer can be adjusted by changing the length of oligomer ligated at the DNA fragment terminus.

Enzyme digestion and bonding of the oligomer having a known sequence by ligation allow the known DNA sequence to be introduced into both sides of the unknown DNA fragments. The introduction of labels into DNA fragments facilitates separation and fractionation of the DNA fragments. After separation and fractionation, five to six or less DNA fragment species are contained in each fraction. Small number of DNA fragment species in a fraction facilitates the DNA sequencing on a selective basis. The fractionated DNA fragments contain the unknown sequences between the known sequences of the ligated oligomers, and can be amplified enzymatically by the extensively used PCR method. The primer containing the oligomer sequence and the additional sequence of up to two to six bases is used for complementary DNA strand extension on a selective basis, where only the specific DNA fragments, to which the additional sequence of the primer can hybridize perfectly, are amplified and read out. Selectivity of the complementary strand extension depends on temperature; the selectivity of the amplification of a special DNA fragment species can be improved by raising reaction temperature over 60 degrees Celsius.

For separation and fractionation of DNA fragments by their sequence difference, it is also possible to use the features of the DNA strand undergoing (hybridization) with DNA strand having complementary sequence. Various types of DNA oligomers are immobilized to the solid surface according to the type, so that the target DNA fragments are hybridized with it. The oligomers have the sequences complementary to the ligated oligomers at the DNA fragments and the additional two to six bass at 3' terminus. When the number of additional bases is two, the number of oligomer species becomes 16 according to the sequence variation of the additional two bases. The oligomers are immobilized on the solid support separated by their additional sequences. Since the target DNAs are double stranded DNAs, they are converted into single stranded DNAs by alkali denaturation, or are partially decomposed from the double strand 3' terminal using such enzyme as exonuclease III, to be made into single strands with the terminal having a form of protruded 5' terminal. Or these DNAs are decomposed from the 5' terminal by λ exonuclease, so that the terminal has a form of protruded 3' terminal; after that, they are hybridized with various types of DNA oligomers immobilized on the solid surface, thereby separating the target DNAs. These target DNAs are partially taken according to the type of said DNA oligomers, to determine the sequence. This process is concurrently implemented for various types of DNA fragments.

To sum up, the oligomer of the known sequence is bonded to the terminal of the DNA fragment obtained by digesting the sample with the restrictive enzyme, and hybridization is made to occur between oligomer and DNA fragment, using the DNA probe chip whose solid surface is fixed with the oligomers having the sequences of combinations of many of the base types (all of them if possible) within the length of the additional base sequence following the known sequence. Furthermore, fractionation of DNA fragments is possible after the hybridization between probes and DNA fragments has been so made that DNA polymerase reaction on the probes will proceed only for the ones completely hybridized with DNA fragments to improve the stability of the hybridomers. Detection is made to find out whether hybridization has occurred in a cell on the solid surface or not, and the terminal sequence of the DNA fragments is known from the address of the cells in which DNA fragments are hold by hybridization; then the individual DNA fragments are removed from the surface to be analyzed as they are. This procedure provides a method of effective analysis of long DNAs mixtures of various DNA fragments, which is applicable to the simultaneous analysis of a great number of the FDNA fragment species having different sequences.

Direct sequencing of long DNAs reaching 1 mega base is very difficult. So the major point of the present invention is to provide a means of completing analyses in a short period of time by parallel sequencing of DNA fragments produced by enzymatic digestion and separated thereafter. When the restriction enzyme recognizing six base is used, the average DNA fragment length is several hundreds. The number of DNA fragment species is about 100. Since the terminal of each DNA fragment has an inherent sequence, separation is possible by recognizing and hybridizing it with the complementary sequence. For stable hybridization it requires the length of complementary sequence of 10 mer or more. The number of oligomer species comprising ten mer is 4^{10}, i.e., about 10^6, so it is not easy to make and prepare all of these oligomer species. Several hundreds of DNA fragment species are present in the digestion product of 1 Mb DNA; therefore, several thousands of oligomers are required to recognize and retain

them. Namely, their identification is possible by using oligomers of five to six mer in their identification region. It is possible to create the DNA probes capable of complementary bonding, having sufficient stability with the target DNA if the DNA probes contain said sequence of 5 - 6 mer for identification and the sequence of the portion digested by restrictive enzyme as well as part or all of the known sequence bonded to the terminal. The target DNA in a solution can be retained and separated according to the base sequences in the identification region of the DNAs which make the complementary bond with DNA probes, by pouring the solution including target DNAs onto the solid surface where these various types of DNA probe are immobilized. The separated DNA undergoes sequencing reaction in parallel and can be analyzed. When the number of DNA fragment species in a mixture is less than 100, the length of the identification region in the probes can be three to four mer.

BRIEF DESCRIPTION OF THE DRAWING

Fig. 1 is a flow diagram representing the DNA analyzing method in one embodiment of the present invention;

Fig. 2 is a structure diagram representing the base sequence of the DNA probe used in one embodiment of the present invention;

Fig. 3 is a schematic illustration to describe the instrument to measure the presence or absence of hybridization reaction used in one embodiment of the present invention;

Fig. 4 is a view showing the fluorescence image pattern comprising DNA fragments hybridized on the probe chip used in one embodiment of the present invention;

Fig. 5 is a schematic illustration to describe other probes (combination of linear probe arrays) used in one embodiment of the present invention;

Fig. 6 is a flow diagram representing the DNA analyzing method in another embodiment of the present invention;

Figs. 7A and 7B are schematic illustrations to describe the labeling of DNA fragments with fluorophore;

Fig. 8 is a schematic illustration to show the DNA fragment separation and separate taking method in another embodiment of the present invention;

Fig. 9 is a schematic illustration to show the DNA fragment amplification procedure using PCR (Polymerase Chain Reaction) process;

Figs. 10A and 10B are schematic illustrations to show selective base sequencing method in another embodiment of the present invention; and

Fig. 11 is a structure diagram representing the base sequence of the primer used in selective base sequencing in another embodiment of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

EMBODIMENT 1

The present embodiment will be described with reference to Figures. The following describes the case of using λ phage DNA having 48,000 bases as one double stranded DNA sample 1 to be analyzed: The λ phage DNA is digested by the restriction enzyme EcoRI (G↓AATTC: restriction enzyme EcoRI recognizes the base sequence GAATTC and cuts between G and A: this description applies to all the following cases), and ethanol precipitation is utilized for purifying decomposed substances. After that, the double stranded oligomer, where the 5' protruded terminal has AATT and double stranded portion has the following sequence, is ligated to this digested portion by using the enzyme, ligase:

TGTAAAACGACGGCCAGTG
*ACATTTTGCTGCCGGTCACTTAA

In Fig. 1, "2" denotes the unknown sequence portion inside the DNA fragment while "*" denotes the fluorophore label or the like. As shown in Fig. 1, the DNA fragment with the 3' terminal sequence of *ACATTTTGCTGCCGGTCACTTAAG... is generated. In the case of λ phage DNA, the number of the digested portions of the restriction enzyme EcoRI is as small as five. The sample is similarly digested with BamH I (G GATCC), Bal I (TGG CCA), BapH I (T CATGA), and the oligomers having known sequence are bonded to the terminal by ligation. The area in vicinity of the 3' terminal of this oligomer is labeled with a fluorophore or a ruthenium complex which is an electroluminescence reagent. It goes without saying that a radioisotope label may be used. In the present invention, Texas Red (trademark of Molecular Probe Co.; having a maximum fluorescence emission wavelength of about 615 nm) was used as the label. After the enzymatic digestion and ligation process, λ DNAs are turned into almost 40 fragments. Exonuclease is added to these reaction products to decompose DNA fragment from the 5' terminal. After they are turned into DNA fragments having protruded 3' terminal, the DNAs are separately taken by the ethanol precipitation and are held in the 1 x TBE buffer of 10 µl (10 microliters).

Probe chip 4 is created by bonding various types of oligonucleotides (DNA oligomer probe 3) at pitches of 0.1 mm on the glass surface measuring 6 mm x 10 mm in the same way as disclosed in the document (Science Vol. 251, pages 766

through 772, 1991). A total of 6,000 separated cells are formed on probe chip 4. Fig. 2 shows the structure of the DNA probe used in the embodiment of the present invention, and the DNA probe has the enzyme recognizing sequence 5 (cuting cites), the universal primer portion of 6 being common sequence among DNA probes and the addition sequence portion 7 to recognize and separate the sample DNA sequence. As illustrated in Fig. 2, oligomers (15 mer - 20 mer) of 1024 species (4^5), which include the cutting cite sequences of 6 mer (the variation of cutting site sequence is four in this case because four types of restriction enzymes are used) and selective sequence portion 7 of 5 mer in the unknown region (x' ... x') of the DNA (the variation of the sequence is $4^5 = 1024$ and this region is used to identify and select DNA fragments), are bonded on the glass surface by changing oligomer species for each cell of probe chip 4 (also referred to as array sensor). The length of the selective sequence portion 7 was assumed as 5 mer; it can be 2 to 6 mer. If the selective sequence portion 7 is too short, various types of samples will be hybridized in one cell. Conversely, if the selective sequence portion 7 is too long, the required oligomer species will increase in number, resulting in difficulties in the fabrication of the said probe chip. When fragments are separated by electrophoresis in advance, the length of the selective sequence portion 7 can be about 2 mer. In order to identify whether complete hybridization at the sequence portion of the 2 mer has been achieved or not, complementary DNA strand extension (this reaction is sensitive to 3' terminal (2 mer)) is performed to improve the stability as the hybridoma, thereby distinguishing it from the hybridoma which was not extended; this method is effective.

The liquid containing DNA is injected on the probe chip, and DNA fragment is hybridized with the oligomer on the probe chip. It is held at the temperature of 37 degrees Celsius for about one hour to allow sufficient hybridization; then it is washed with 1 x TBE buffer solution. The state of hybridization on the probe chip is measured by the instrumentation system given in Fig. 3. The probe chip 10 is irradiated by laser 8 (YAG laser, oscillation wavelength of 532 nm and output of 10 mW) through collimator lens 9, and the fluorescence image on probe chip 10 is observed by highly sensitive 2-dimensional camera 11 provided with filter 13. The similar measurement is possible by narrowing the laser beam and using the laser microscope which scans the probe chip range. In Fig. 3. "12" denotes data processor. Fig. 4 shows an example of the observed fluorescence image schematically. Various DNA probes are immobilized on cells in the probe array 14. Fluorescent light is observed only from cell 16 where the target DNA fragment is held. As each cell holds one probe species, the terminal base sequence can be identified from the cell position where fluorescence is emitted. It is also possible to fractionate and use the DNA trapped in the cell. Fractionation of the DNA from each cell can be performed based on the invention disclosed in Japanese Patent Application Laid-open NO.5-236997 (Application No. 4-42829 "Polynucleotide Capturing Tip": the corresponding patent application in U.S. is U.S. Patent Application Ser. NO.08/021,667).

When fractionation of a large number of DNA copies is desired, the magnetic beads (fine iron particles whose surface are coated with the organic substance to permit DNA bonding) on the surface of which retains DNA oligomer having the same sequence as that on each cell are put into the sample tube, thereby causing hybridization of DNA with the oligomer on the beads and separation is carried out by taking out the beads with DNA using a magnet. This procedure should be repeated.

Or fractionation method as shown in Fig. 5 is also available. That uses plural of linear DNA probe array comprising thin plates. Each probe array has many cells holding different oligomer probes, respectively, however, only one cell per array is used to fractionate DNA at a time. Each probe array can be separated for fractionating the DNA hold in the cell. For this purpose, one or more of these chips are prepared and are laid out superimposed by shifting the position of more than one liner DNA probe array 17 to the direction where separated cells are laid out. The flow channel 18 for sample solution is formed so that the separated cell, where the probe corresponding to the DNA fragments to be fractionated on the specific line is fixed (DNA probe cell 19 to capture the target DNA), will be arranged in one line. Hybridization is made by pouring the sample solution into flow channel 18 and more than one target DNA can be in sequence captured, thereby ensuring effective fractionation operation.

The probe sequences to be used are selected by shifting the position of the chips, as shown in Fig. 5. After the hybridization and capturing of DNA fragments on the chips, each chip is separated and specific DNA fragments are recovered separately. Different DNA fragments can be captured by shifting chips variously.

It is also possible to use the following procedure: Various types of probe rods with DNA oligomer probes mounted on the slender bar tip are prepared to select the probe rods having the required sequence. These rods are bundled and put into the sample solution, thereby causing hybridization for separate taking of the DNA. This means that the probe holding medium such as rods which can be separated each other can be

used in place of the sheet containing cells holding various probes on them. In this case, for selecting probe rods where the probe corresponding to the DNA fragment are fixed, the two-dimensional probe array composed by more than one linear DNA probe array 17 arranged in two dimensions is analyzed and inspected in advance, using the device as described with reference to Figs. 3 and 4 before fractionation of the DNA fragment. The DNA with the known terminal sequence can be singly separated with this method. When the DNA length is 1 to 2 kb or less, the number of sample DNA copies can be amplified by PCP (Polymerase Chain Reaction); therefore, DNA can be separated and amplified directly from probe chips 4 and 10, and probe arrays 14 and 17, or PCR amplification can be performed directly on the probe chips 4 and 10, and probe arrays 14 and 17 by installing a barrier to partition the sample holding cells from other cells on these chips and arrays. For the PCR primer, it is possible to use complementary oligomer on both terminals including the known sequence. When the probe comprising the DNA oligomers on the probe chip is hybridized with the sample DNA at its 3' terminal side, and complementary DNA strand extension is possible from the probe oligomer 3' terminal side along the sample DNA strand, then the complementary strand is formed on the probe chip firstly and can be used as a template for PCR amplification. In this case, the same strand as that of the sample DNA is taken up by thermal denaturation, and is used for analysis such as sequencing.

DNA sequencing reaction with the fractionated DNA fragment according to the normal procedure, and DNA sequence is determined by gel electrophoresis or other means. Since the sequence portion common to the oligomers ligated to the DNA can be used as sequencing primer, it eliminates the need of preparing many types of primers. Sequencing reaction and operation (by gel electrophoresis) can be performed for many samples at one time. If long DNAs are digested and separated by the present invention, these separated samples can be sequenced at one time. If fragmentation of the sample by restriction enzyme is repeated several times by the enzymes with different combinations, all the sequences can be determined exhaustively. Furthermore, if one sequencing is not sufficient to sequence the required length, the new primer is synthesized in the vicinity of the terminal of the DNA sequence read out, and the sequence of a stiller longer portion is determined.

In the embodiment discussed above, the target DNA is screened and separated by the 5-mer base sequence following the portion digested by enzyme. When there are only a few types of the target DNA fragments, screening and separation are possible by two to four-mer base sequence. When there are many types of the fragments, they are fractionated roughly into 10 fractions by gel electrophoresis; then the oligomer having a short sequence for screening for each group can be used for analysis and inspection.

In the embodiment discussed above, the DNA made to have single strands by exonuclease is obtained by hybridization. It is also possible to obtain the target DNA by hybridization after single strands have been obtained by thermal denaturation. Furthermore, the DNA fragments immobilized on the glass can be obtained by complementary DNA strand extension reaction and thermal denaturation, and can be used for sequencing analysis. Furthermore, PCR amplification can be made by the primer (DNA oligomer probe) and immobilized on the solid surface under this condition, as well as the universal primer, thereby increasing the number of copies to be used for analysis.

EMBODIMENT 2

Another embodiment of the present invention will be described with reference to Figs. 6 through 11. Fig. 6 represents the flow of the new DNA sequencing method, while Figs. 7A and 7B through 11 provide detailed illustrations of the techniques in each process. The DNA 101 to be sequenced and analyzed is digested by the restriction enzyme (Hind III) 102 which recognizes and digests specific sequences. This operation generates DNA fragment group 103 with its terminal base sequenced.

Whenever required, this DNA fragment terminal base is labeled with fluorophore 104 in the method shown in Figs. 7A and 7B. (1) in Fig. 7A denotes the method of incorporating the monomer 112 of DNA labeled with fluorophore into the terminal of the DNA fragment 103, using the DNA synthetic enzyme. FITC (fluoresceine isothiocyanate having an emission wavelength of 525 nm) is used as the label of DNA monomer 112. Fig. 7B shows the method of bonding the short DNA 113 having the protrusion which can be bonded with the protruded portion of one strand of the terminal digested by restriction enzyme 102, by means of DNA bonding enzyme such as ligase. The short DNA fragment 113 is labeled with fluorophore 104 in advance through amino residue or the like. Texas Red (sulforhodamine 101; having an emission wavelength of 613 nm) is used as the label.

The generated DNA fragment group is separated and taken out by electrophoresis using gel 114 whenever required, as shown in Fig. 8. According to the present embodiment, acrylamide gel of 8 percent T (total acrylamide concentration) and 3 percent C (concentration of chemical agent for cross link) filled in the glass tube 115 having an

inner diameter of 2 mm was used in the separating portion. Shorter DNA fragments 120 move faster, so they are taken separately from the shorter fragments into the container 119 sequentially. Laser 121 having irradiated on a specified position in the glass tube excites the labeled fluorophore 104, when the DNA fragment has passed by. This make it possible to estimate the length of the DNA fragments taken separately by monitoring the fluorescent light emitted from the moving DNA, as well as the number of the mixed fragments. The fractionation of DNA fragments utilizes the difference of the molecular weights of DNA fragments or the specificity of the base sequence. When the PCR operation is to be performed, this operation of fractionation provides effective pretreatment. PCR is performed in order to get the amount required for the base sequencing operation, whenever required.

PCR provides a method of amplifying the copy number of DNA fragments 103 enzymicatically, as shown in Fig. 9. The DNA fragments 106 digested by restriction enzyme (Hind III) 102 and fractionated are dissociated to single stranded DNA 122 by temperature rise. Since the sequence on each 3' terminal side is the known sequence, DNA strand (primer) 123 which can be bonded thereto is hybridized by reducing the temperature. With primer 123 as the starting point, each complementary strand is synthesized, and the copy 108 of original DNA is generated. The DNA strand generated is dissociated into one strand by raising the temperature again. When this cycle is repeated n-times, the DNA copy number is amplified to 2 raised to n-th power times.

Five to six types of DNA fragments 107 are obtained in a great amount by such method; of them, only one type is selected by the selective base sequencing method shown in Figs. 10A and 10B, resulting in determination of the base sequence. DNA fragment group 107 including five to six fragment species as shown in Fig. 6. The fragments have the known sequence portion comprising the portion digested by restriction enzyme on both terminals and oligomer connected thereto, and the unknown sequence portion between them.

This DNA is made into the single stranded DNA 109 or 111 as shown in Figs. 10A and 10B; then this is bonded with the primer 110 comprising the base sequence complementary with the base sequence, which is composed of said known sequence portion and two bases on the 3' terminal side of its ensuing unknown sequence portion. The two bases on the 3' terminal side of the DNA fragment pertain to the unknown sequence, and the type of the base sequence varies depending on the DNA fragment species. Therefore, the type of the primer which perfectly hybridizes to each of them varies depending on the DNA fragment species. Fig. 10A shows the case of perfect hybridization with DNA fragment 109, whereas Fig. 10B shows the case of a not perfect hybridization with DNA fragment 111. Namely, Fig. 10A shows the case where complementary DNA strand extension reaction takes place, but Fig. 10B shows the case where complementary DNA strand extension reaction does not proceed. Thus, of DNA segment group 107 comprising five to six species, only one type of DNA fragments (may be more than one species depending on the case, where selective sequence must be increased from 2 to 3) can selectively determine the base sequence.

Regarding the structure of the primer required for this selective base sequencing operation, Fig. 11 shows the case there are two DNA fragment species in a group which have the same sequence of the 3' terminal of the unknown sequence portion, for example. The portion 128 in the primer, which is hybridized with the known sequence portion of the 3' terminal of the DNA fragment, comprises the portion 126, which is hybridized with the known sequence newly bonded by ligation or the like, and portion 127 which is hybridized with restriction enzyme recognizing sequence. The 3' terminal side of the primer which follows the known sequence region is bonded to the unknown sequence portion of DNA fragment, so four types of the first base and four types of the second base (4 x 4 = 16), namely, sixteen types, are prepared to cover all cases. This is used not only for base sequencing but also for amplification of only the specific fragments during PCR amplification.

Since hybridization between the said primer and DNA fragment is attributable to hydrogen bonding, hybridization is stabilized at the 3' terminal portion only when these hybridizations have perfectly matched sequence, if hybridization temperature is raised. Therefore, if hybridization reaction temperature (annealing temperature) is set to 60 degrees Celsius or more, more effective selection will be made, resulting enhanced selectivity.

Making DNA have the known sequence portion and the unknown sequence portion and using the selection primer according to the above principle, we performed base sequencing operation normally used as described below.

In the presence of 10 mM Tris-HCl, pH 8.5, 6 mM $MgCl_2$, 1 pmol of primer, 0.45 pmol of sample DNA and 1 unit of AmpliTaq $_R$ DNA polymerase, which is thermostable enzyme, are mixed together to get 15 μl. The ddATP as a terminator is added to dATP, dCTP, dGTP and dTTP as the DNA solution in the sample tube with 0.5 ml capacity to get a total of 1 μl in advance; then it is dispensed into the A-reaction tube. The A-reaction tube thus prepared, C-reaction tube containing ddCTP instead of ddATP, G-reaction tube containing ddGTP

and T-reaction tube containing ddTTP are prepared in advance. 3.5 µl of the mixed solution of the primer and sample is dispensed into each of the A, C, G and T tubes. Furthermore, 1 to 2 drops of mineral oil are put into them, and set them onto DNA thermal cycler. For cycle reaction conditions, temperatures of 95 degrees Celsius for 30 seconds and 72 degrees Celsius for one minute are repeated in the case of the cycle time being 1 to 15 cycles. In the case of 15 to 30, temperatures of 95 degrees Celsius for 30 seconds and 72 degrees Celsius for one minute are repeated. DNA complementary strand extension reaction is carried out by this cyclic reaction, thereby generating the DNA fragments where 3' terminal are A from A reaction, C from C reaction and so on. After completion of cyclic reaction, 2 µl of formaldehyde, reaction stop solution, is added, and the reaction mixtures are loaded on the gel which is set to the fluorescent DNA sequencer. 1,400-volt voltage is applied to 40 cm gel to separate the DNA fragment and to determine the base sequence.

In this case, since hybridization reaction temperature (annealing temperature) is set at 60 degrees Celsius or more, so Taq cycle sequencing was performed using Taq polymerase thermostable enzyme. Other thermostable enzymes can be used. Depending on number of fragment species in a sample, selective hybridization may be performed without setting the hybridization reaction temperature (annealing temperature) to 60 degrees Celsius or more. In such cases, base sequencing operation can be performed using the polymerase which is not thermostable enzyme. Selective base sequencing operation can be performed using the terminator sequencing method where reaction is carried out by labeling the terminator (ddNTP) if the primer is not labeled.

In the above drawing, numeral 105 denotes the DNA fragment group digested by the restriction enzyme and labeled with fluorophore, numeral 116 the upper buffer cell, numeral 117 the lower buffer cell, numeral 118 electrode, numeral 124 the known sequence on the 5' terminal side of single stranded DNA fragment, numeral 125 the known sequence on the 3' terminal side of single stranded DNA fragment, and numeral 129 the known sequence portion and two- to six-base portion bonded to the unknown sequence portion bonded to the 3' terminal side following it.

The present invention permits simple base sequencing without requiring any specific facility such as P2 facility for cloning of target DNA. It does not require culturing, so base sequencing operation including sample preparation can be completed in a shorter period of time. Furthermore, this method allows all operations to be done in test tubes, encouraging automation of the operation process. According to the present invention, furthermore, a long DNA is digested and separated to identify the sequence at both terminals of the digested and separated DNA, and each fragment can be separated by using it. This method further allows DNA analysis in a short time by simultaneous analysis of the separated fragments. For example, if the DNA of 1 mega base is separated into 500 fragments having an average strand length of about 2 kB and is analyzed in parallel, then terminal sequence search and separation can be completed in two or three days; sequencing can be completed in three or four days by using gel capillary array and primer walking. So analysis is completed in about one week, thereby allowing analysis to be made 100 times faster than the conventional method.

Claims

1. A DNA analyzing method comprising:
   i) a process step of forming DNA fragments by digesting a DNA strand,
   ii) a process step of recognizing the base sequences in the vicinity of the terminals of DNA fragments, and fractionating or collecting the DNA fragments separately, and
   iii) a process step of sequencing each of DNA fragments recognized and fractionated or separately collected in process step ii).

2. A DNA analyzing method comprising
   I) a process step of digesting the double stranded DNA,
   II) a process step of bonding a DNA oligomer having a known base sequence to the digested portion of the DNA,
   III) a process step of recognizing the base sequence of the DNA which follows the known base sequence at the terminal of the DNA after having carried through process step II), and fractionating or collecting the digested DNA fragments separately according to the difference of the base sequence in the vicinity of the terminals of the DNA fragments, and
   IV) a process step of determining the base sequence of the DNA fractionated or separately collected in process step III.

3. A DNA analyzing method according to Claim 1, wherein, in process step ii), oligomers having different base sequences are each fixed separately on an array sensor for each base sequence of oligomer, and the base sequences in the vicinity of the terminals of said DNA fragments are recognized with the oligomer fixed array sensor.

4. A DNA analyzing method according to Claim 3, wherein, said array sensor in the process ii) comprises a combination of a plurality of movable line sensors in which each of oligomers having various sequence is immobilized in a cell arrayed linearly.

5. A DNA analyzing method according to any of the Claims 1, 3 and 4, which includes, between said process steps i) and ii), another process step i') of bonding the terminal of the DNA fragment with DNAs having known base sequences which are linked with an agent selected from the group consisting of fluorophore, electroluminescence reagent and radioisotope label.

6. A DNA analyzing method according to Claim 5, which, in said process step ii), allows recognition of the base sequence in the vicinity of the DNA fragment terminal by hybridization between the vicinity of said DNA fragment including, at least, part of the known base sequence and oligomer on the array sensor.

7. A DNA analyzing method according to Claim 6, wherein, in said process step ii), the base sequence of the oligomer on the array sensor is complementary to the base sequence comprising:

a) base sequence of, at least, part of the labeled DNA having the known base sequence

b) base sequence of the recognition sequence recognized by the restriction enzyme used to digest the DNA in process i), and

c) the base sequence having at least two-base length for identifying DNA fragments, following the enzyme recognizing sequence.

8. A DNA analyzing method according to Claim 6, which, in said process ii), allows detection of the presence or absence of hybridization between said DNA fragment and oligomer on the array sensor, recognition of the base sequence at the terminal of the DNA fragment according to the result, and fractionation, separate taking or analysis of said DNA fragment on the array sensor.

9. A DNA analyzing method according to Claim 7, wherein said base sequence c) has a length of 2 to 6 mer.

10. A DNA analyzing method according to Claim 2, which includes between said process steps II) and III), another process step II') of converting at least the portion in the vicinity of the terminal of said double stranded DNA into a single stranded form.

11. A DNA analyzing method according to Claim 2 or Claim 10, wherein, in said process step III), more than one type of oligomer having different base sequences is separated for each base sequence, and the base sequence of the DNA which follows the known base sequence at the terminals of said DNA fragments is recognized on a fixed array sensor.

12. A DNA analyzing method comprising A) a process step of digesting double stranded DNA, B) a process step of bonding an oligomer having a known base sequence to the digested portion of the DNA, and C) a process step of recognizing the base sequence portion of the DNA following the known base sequence, and synthesizing the DNA complementary to the specific fragment.

13. A DNA analyzing method according to Claim 3 or Claim 11 wherein, in said process step ii) or III), the base sequence in the vicinity of the terminal is recognized through complementary DNA strand extension.

FIG. 1

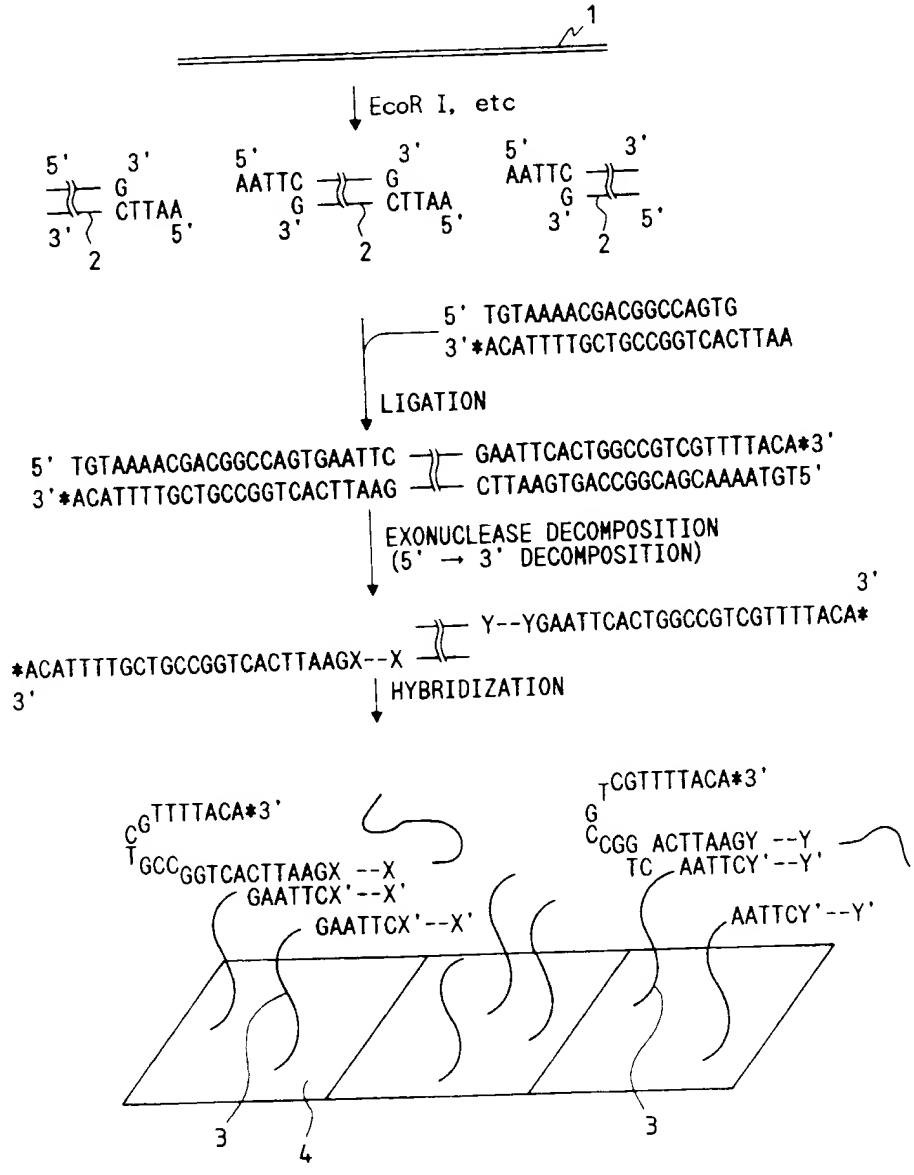

1
EcoR I, etc
5' 3'
G
CTTAA
3' 5'
2
5' 3'
AATTC
G
CTTAA
3' 5'
2
5' 3'
AATTC
G
3' 5'
2
5' TGTAAAACGACGGCCAGTG
3'*ACATTTTGCTGCCGGTCACTTAA
LIGATION
5' TGTAAAACGACGGCCAGTGAATTC
3'*ACATTTTGCTGCCGGTCACTTAAG
GAATTCACTGGCCGTCGTTTTACA*3'
CTTAAGTGACCGGCAGCAAAATGT5'
EXONUCLEASE DECOMPOSITION
(5' → 3' DECOMPOSITION)
3'
Y--YGAATTCACTGGCCGTCGTTTTACA*
*ACATTTTGCTGCCGGTCACTTAAGX--X
3'
HYBRIDIZATION
CG TTTTACA*3'
T
GCC GGTCACTTAAGX --X
GAATTCX'--X'
GAATTCX'--X'
T CGTTTTACA*3'
G
C CGG ACTTAAGY --Y
TC AATTCY'--Y'
AATTCY'--Y'
3
4
3

FIG. 2

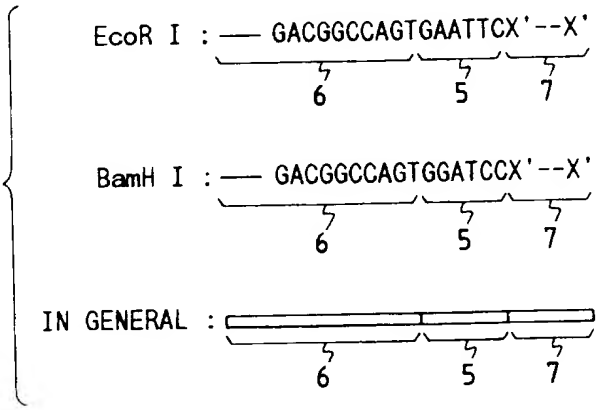

EcoR I : — GACGGCCAGTGAATTCX'--X'
6
5
7
BamH I : — GACGGCCAGTGGATCCX'--X'
6
5
7
IN GENERAL :
6
5
7

FIG. 3

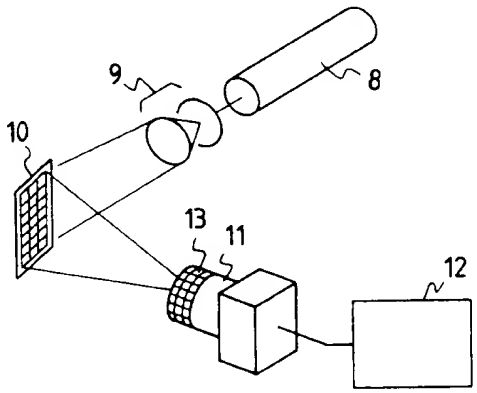

9
8
10
13
11
12

FIG. 4

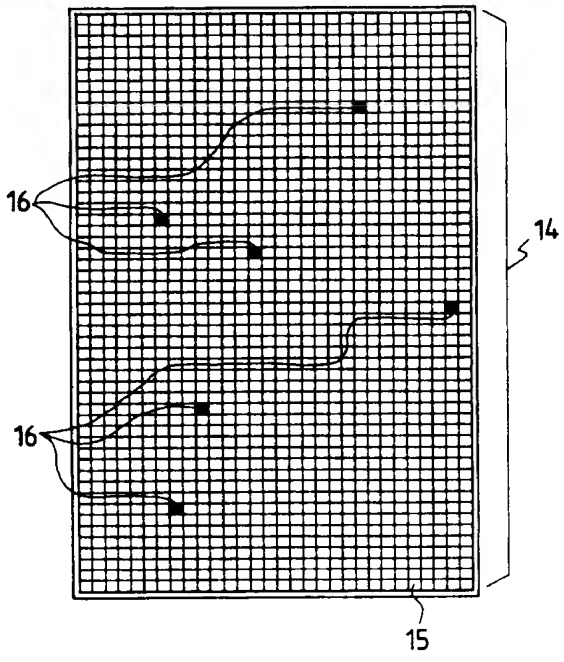


FIG. 5

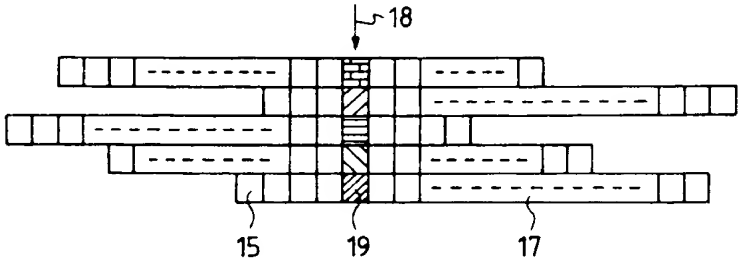


FIG. 6

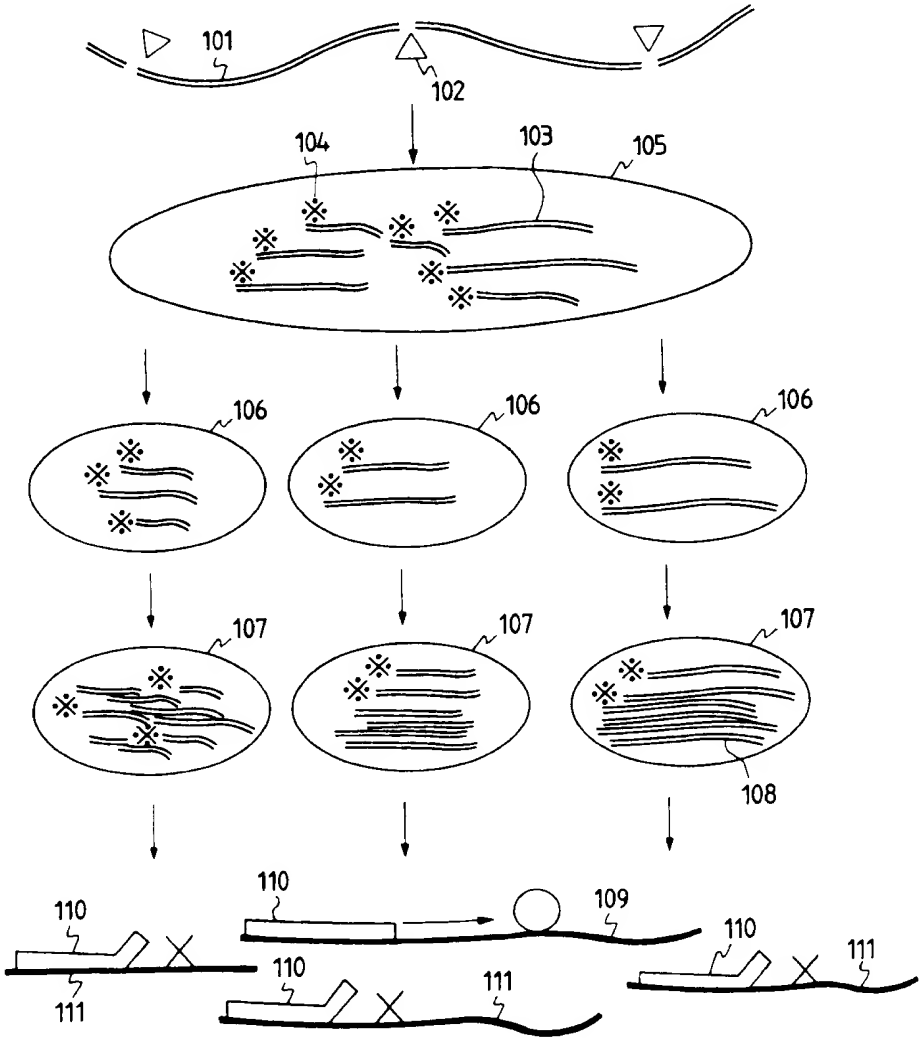

101
102
104
103
105
106
106
106
107
107
107
108
110
110
109
110
111
110
111
111

FIG. 7A

FIG. 7B

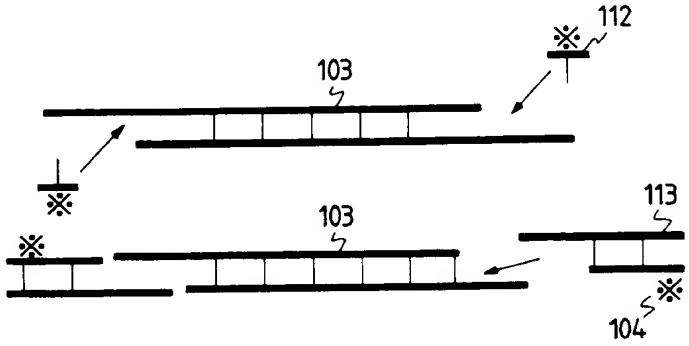


FIG. 8

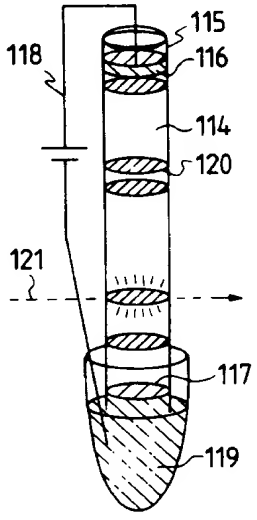


FIG. 9

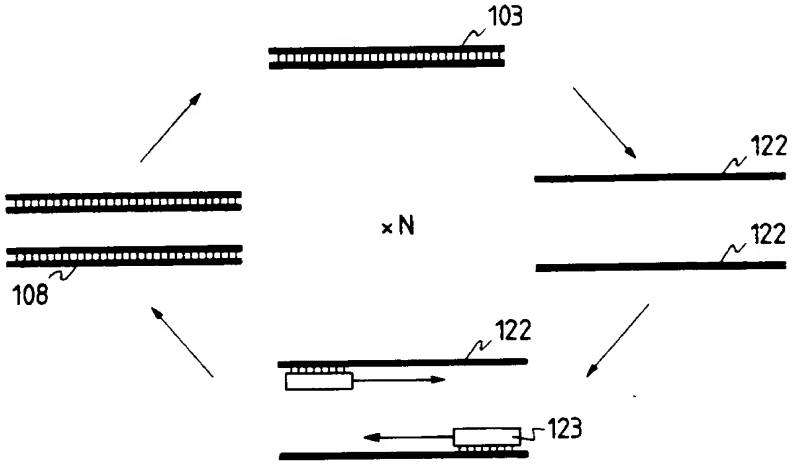


FIG. 10A

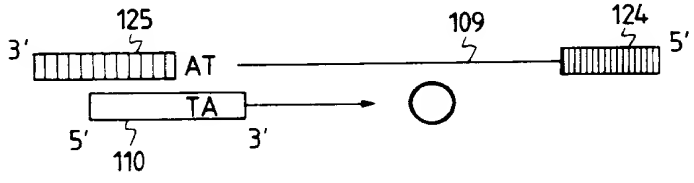


FIG. 10B

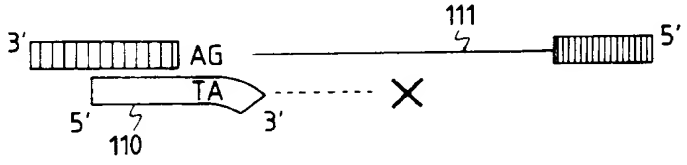
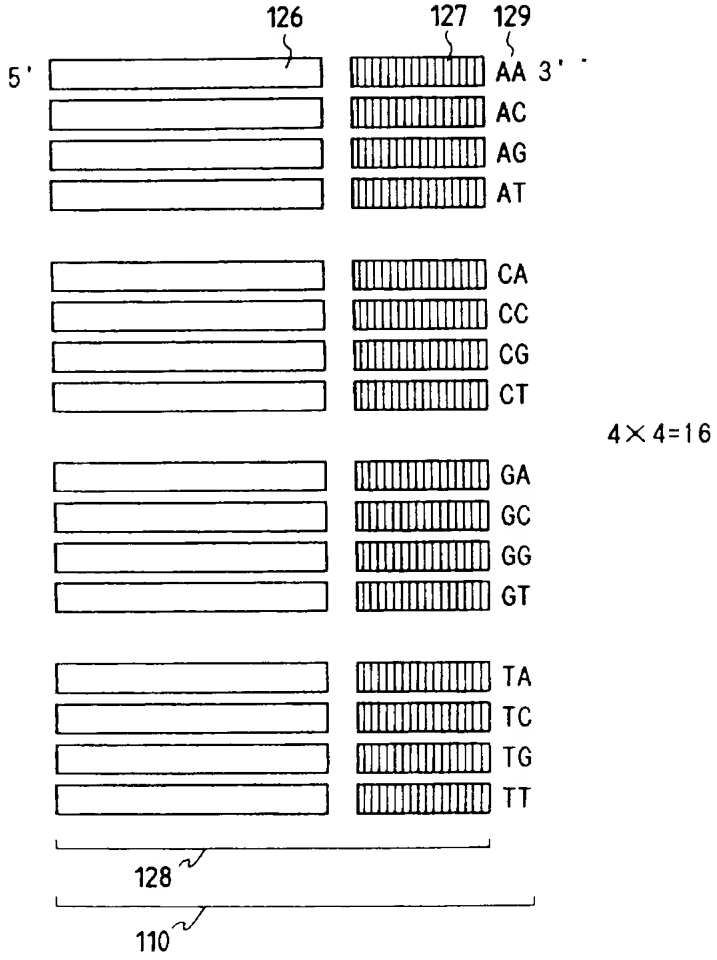


FIG. 11


126
127
129
5'
AA 3'
AC
AG
AT
CA
CC
CG
CT
4×4=16
GA
GC
GG
GT
TA
TC
TG
TT
128
110